İndi suallara cavab vermə zamanı qısa məzmun çıxartmanın istifadəsinə baxaq.

Qısa məzmun çıxartmanın məqsədi verilmiş mətnin istifadəçi və ya xüsusi

istifadəçi ehtiyacı üçün ixtisar olunmuş versiyasının yaradılmasıdır.

Buna misal olaraq hər hansı növ sənədin və ya məqalənin xülasəsinin hazırlanmasını,

elektron poçt yazışmalarının və ya görüşlərin qısa məzmununun yaradılmasını,

mətnin sadələşdirilməsini və s.-ni misal göstərmək olar.

Biz cümlələrin sıxılmasından sonra danışacağıq.

Qısa məzmun çıxartma haqqında iki yolla danışa bilərik.

Bir sənədin qısa məzmununun çıxardılması.

Biz verilmiş sənəd üçün xülasə və ya konspekt, bəlkə başlıq, bir sözlə

desək sənədin çox qısa məzmununu yaradacağıq.

Bir neçə sənədin qısa məzmununu çıxardığımız zaman, bizə sənədlər qrupu verilir,

yəqin ki bu sənədlər bir-birilə əlaqəlidir və bizim işimiz həmin sənədlərin

məğzini və ya mahiyyətini yaratmaqdır. Əgər bizdə bir hadisə ilə bağlı xəbərlər

toplusu varsa, onda həmin hadisənin sənədlərdən informasiya götürməklə qısa məzmununu

çıxarda bilərik. Və bu hər hansı mövzu və ya sual ətrafında fokuslanmış istənilən

veb səhifələr toplusu üçün də doğru ola bilər.

Indi biz ümumi qısa məzmun çıxartma (generic summarization) ilə

sorğu-əsaslı qısa məzmun çıxartmanı (query-focused summarization) fərqləndirəcəyik.

Ümumi qısa məzmun çıxartma bizim indiyə qədər danışdıqlarımız idi.

Bizim bir sənədimiz və ya hər hansı sənədlər dəstimiz var, və bizim işimiz

sadəcə qısa məzmun yaratmaqdır.

Sorğu-əsaslı qısa məzmun çıxartma isə istidaçinin sorğu vasitəsilə göndərdiyi

informasiyaya əsaslanaraq sənədin xülasəsinin çıxarılmasıdır. Beləlikə siz

sorğu-əsaslı qısa məzmun çıxartmanı digər tip kompleks sual cavablandırma kimi

başa düşə bilərsiz. Biz sualı istifadəçinin ehtiyac duyduğu cavabı yaratmaq üçün

içərisində informasiya olan sənədin qısa məzmununu çıxartmaqla cavablandırırıq.

Biz hamımız sorğu-əsaslı xülasələrlə tanışıq, çünki bu axtarış sistemlərinin

səhifə haqqında informasiyanı sizə göstərmək üçün istifadə etdiyi metoddur.

Məsələn, Google sizə hər bir səhifə üçün 156 simvol - təqribi 26 sözdən ibarət

səhifənin xülasəsini verəcək. Əgər mən Google'a "what is die brücke" sualını

versəm, deyək ki, bu üç veb ünvanının əldə edəcəm. Biz veb ünvanının adına,

veb ünvanının ünvan göstərgəsinə, və sonda bir qədər mətnə malik olacağıq.

Və o mətn - mətn parçası istifadçiyə səhifədə nə olduğunu başa düşməyə kömək edən

bir növ səhifənin qısa məzmunudur. Beləliklə, biz tək

sənədlərdən qısa məzmun çıxartmanı mətn parçaları kimi şeylər yaratmaq

kimi başa düşə bilərik. Və biz veb axtarışının nəticəsi olan mətn parçalarından fərqli

olaraq daha ümumi məsələrdən danışacağıq, ancaq bu mətn parçalarını simvolik nümumə

kimi başa düşə bilərsiniz. Mətn parçası tək sənəddən informasiya götürməklə sualın

cavabı olurdusa, biz tez-tez bir neçə sənəddən cavab çıxartmaq istəyirik. Kompleks

suallara müxtəlif sənədlərdən qısa məzmun çıxartmaqla cavab yaratdığımız zaman, biz

bu sənədlərdəki informasiyanı birləşdirən əlaqəli cavab istəyirik. Bu müxtəlif

sənədlərdən istifadə etməklə suala cavab vermədir.

Biz qısa məzmun çıxartma prosesini iki yolla edə bilərik.

Çıxarış yollu qısa məzmun hazırlayarkən (extractive summarization) biz

sənəddə olan xüsusi sözləri və ifadələri götürürük və qısa məzmunu

həmin sözlər və ifadələrdən istifadə etməklə hazırlayırıq. Və Google kimi axtarış

sistemindən bizə qayıdan mətn parçası çıxarış yollu qısa məzmun hazırlanmasına

nümunədir.

Bundan fərqli olaraq abstract qısa məzmunda (abstractive summary) isə

verilmiş mətnin xülasəsi üçün biz mətndəkilərdən

fərqli öz sözlərimizi yaradırıq. Və biz bügün tamamilə çıxarış yolu qısa məzmun

hazırlamaq haqqında danışacağıq. Abstrakt qısa məzmun hazırlama önəmli elmi-araşdırma

məqsədidir, ancaq çox çətindir. Və biz sənəddən (veb səhifə və ya başqa bir şey)

qısa məzmun çıxartmadan danışdığımız zaman, düşündüyümüz tək şey məzmun yaradanda

bizim əsaslarımızın (ana xəttimizin) nə olacağıdır. Və yaxşı yazıçılar öz

ideyalarını tez-tez başlanğıcda, yəni başlıqda və ya birinci cümlədə yerləşdirirlər.

Beləliklə, istənilən növ qısa məzmun çıxartma alqoritmini yerinə yetirdiyimiz zaman

bizim sadə əsasımız (simple baseline) məhz birinci cümləni nəzərə almaqdır. Məsələn,

bu Google sorğusunda: "what is Die Brücke", əgər biz çıxışı birinci raslantıdan

götürsək o zaman biz bu cümləyə baxacağıq "Die Brücke was a

group of German expressionist artists formed in Dresden" və s.

Mətn parçalarının gəldiyi səhifəyə baxaq. Və bizə görə bilərik ki, bu mətn paçası

sadəcə səhifənin ilk 156 simvolundan ibarətdir. Bəzən, ən yaxşı mətn parçası

sadəcə başlanğıcdır. Biz bunu əsas (ana xətt) kimi istifadə etməkdən əlavə

görəcəyik ki, bu müxtəlif növ məzmun çıxartma alqoritmlərində istifadə edə

biləcəyimiz bir xususiyyətdir. Biz verilmiş sənəddən qısa xülasələr və

hətta başlıqlar çıxartmaqla təbii dil əmələ gətirmə məsələsini təqdim etdik.

Və xüsusi suala cavab vermək üçün sorğu-əsaslı qısa məzmun çıxartma metodundan

danışdıq. Biz istifadəçi tərəfindən verilən suala cavab vermək üçün sənədin

qısa məzmununu xüsusi dizayn edilmiş şəkildə yaratdıq. Və biz tək və müxtəlif

sənədlərdən qısa məzmun çıxartmaqdan danışdıq. Və biz növbəti leksiyalarda

bu məsələləri nece yerinə yetirmək lazım olduğunun təfsilatlarına baxacağıq.